On Wasserstein geometry of the space of Gaussian measures

Asuka TAKATSU

Abstract

The space of Gaussian measures on a Euclidean space is geodesically convex in the L^2-Wasserstein space. This space is a finite dimensional manifold since Gaussian measures are parameterized by means and covariance matrices. By restricting to the space of Gaussian measures inside the L^2-Wasserstein space, we manage to provide detailed descriptions of the L^2-Wasserstein geometry from a Riemannian geometric viewpoint. We first construct a Riemannian metric which induces the L^2-Wasserstein distance. Then we obtain a formula for the sectional curvatures of the space of Gaussian measures, which is written out in terms of the eigenvalues of the covariance matrix.

1. *Introduction*

In this paper, we give a formula for sectional curvatures of the space of Gaussian measures on $\mathbb{R}^d$ with the L^2-Wasserstein metric. Let $N(m, V)$ be the Gaussian measure with mean m and covariance matrix V. Namely m is a vector in $\mathbb{R}^d$ and V is a symmetric positive definite matrix of size d and its Radon-Nikodym derivative is given by

$$\frac{dN(m,V)}{dx} = \frac{1}{\sqrt{\det(2\pi V)}} \exp\left[-\frac{1}{2}\langle x-m, V^{-1}(x-m)\rangle\right].$$

We denote by $\mathcal{N}^d$ the space of Gaussian measures on $\mathbb{R}^d$. Since Gaussian measures depend only on the mean m and the covariance matrix V, the space $\mathcal{N}^d$ is identified with $\mathbb{R}^d \times \mathrm{Sym}^+(d,\mathbb{R})$, where $\mathrm{Sym}^+(d,\mathbb{R})$ is the space of symmetric positive definite matrices of size d.

The L^2-Wasserstein space is the subspace of probability measures equipped with a certain distance. Let $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ be the set of absolutely continuous probability measures with finite second moments on $\mathbb{R}^d$. Then $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ is a geodesic space and all geodesics are given by push-forward measure. In view of these facts, Otto [**13**] regarded $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ as an infinite dimensional formal Riemannian manifold and analyzed the porous medium equations as gradient flows on $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$.

A foundation for this framework was carefully laid out by Carrillo-McCann-Villani [**3**]. They introduced the new space, *Riemannian length space*. In short, this space is a length space which has an exponential map defined on some tangent vector space with a metric. They proved that $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ is a Riemannian length space and its metric induces the L^2-Wasserstein distance. We call this metric the L^2-Wasserstein metric.

McCann showed in [**9**] that varying the mean is equivalent to a Euclidean translation and $\mathcal{N}_0^d$ is a geodesically convex subspace of $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$. When we consider the L^2-Wasserstein geometry on $\mathcal{N}^d$, it suffices to consider the geometry on covariance matrix variations. We use $\mathcal{N}_0^d$ for the set of all Gaussian measures with mean 0. We denote by $N(V)$ the Gaussian measure with mean 0, and covariance matrix V.

2000 *Mathematics Subject Classification* 60D05(primary), 28A33.

This work was partially supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

In the Riemannian length space, if a geodesic from $N(V)$ with direction ψ passes thorough $N(U)$, then the gradient of ψ is given as a linear map associated with a symmetric matrix depending only on V, U. Thus the tangent space at each point can be regarded as the space of symmetric matrices $\mathrm{Sym}(d, \mathbb{R})$. This identification coincides with the viewpoint from the differential structure; since $\mathcal{N}_0^d$ is identified with $\mathrm{Sym}^+(d, \mathbb{R})$, which is an open subset of $\mathrm{Sym}(d, \mathbb{R})$, we can consider the tangent space of $\mathcal{N}_0^d$ as $\mathrm{Sym}(d, \mathbb{R})$.

These observations enable us to obtain a formula for the sectional curvature of $\mathcal{N}_0^d$ by restricting to a geodesically convex submanifold $\mathcal{N}_0^d$ of $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$.

THEOREM 1.1. *For an orthogonal matrix P and positive numbers $\{\lambda_i\}_{i=1}^d$, we set $V = P\mathrm{diag}[\lambda_1, \ldots, \lambda_d]{}^{T}P$, where ${}^{T}P$ is the transpose matrix of P. Then we can consider the tangent space to $\mathcal{N}_0^d$ at $N(V)$ spanned by*

$$\left\{ e_+ = \frac{P\,(E_{11} + E_{dd})\,{}^{T}P}{\sqrt{\lambda_1 + \lambda_d}},\ e_{ij} = \frac{P\,(E_{ii} - E_{jj})\,{}^{T}P}{\sqrt{\lambda_i + \lambda_j}},\ f_{ij} = \frac{P\,(E_{ij} + E_{ji})\,{}^{T}P}{\sqrt{\lambda_i + \lambda_j}} \right\}_{1 \leq i < j \leq d},$$

where E_{ij} is an (i, j)-matrix unit, whose (i, j)-component is 1, 0 elsewhere. Then we obtain the following expressions of the sectional curvatures with respect to the vectors:

$$K(e_+, e_{ij}) = 0 \tag{1}$$
$$K(e_+, f_{1d}) = 0 \tag{2}$$
$$K(e_+, f_{ij}) = \frac{3\lambda_i\lambda_j}{(\lambda_i + \lambda_j)^2(\lambda_1 + \lambda_d)} \qquad (i = 1 \text{ or } j = d) \tag{3}$$
$$K(e_+, f_{kl}) = 0 \qquad (1 < k < l < d) \tag{4}$$
$$K(e_{ij}, e_{kl}) = 0 \tag{5}$$
$$K(e_{ij}, f_{kl}) = 0 \qquad (\{i, j\} \cap \{k, l\} = \emptyset) \tag{6}$$
$$K(e_{ik}, f_{ij}) = \frac{3\lambda_i\lambda_j}{(\lambda_i + \lambda_j)^2(\lambda_i + \lambda_k)} \qquad (j \neq k) \tag{7}$$
$$K(e_{ij}, f_{ij}) = \frac{12\lambda_i\lambda_j}{(\lambda_i + \lambda_j)^3} \tag{8}$$
$$K(f_{ij}, f_{kl}) = 0 \qquad (\{i, j\} \cap \{k, l\} = \emptyset) \tag{9}$$
$$K(f_{ij}, f_{ik}) = \frac{3\lambda_j\lambda_k}{(\lambda_i + \lambda_j)(\lambda_j + \lambda_k)(\lambda_k + \lambda_i)} \qquad (j \neq k). \tag{10}$$

The formula coincides with a formal expressions of sectional curvatures of $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ given by Otto [**13**]. This formula shows that the sectional curvature of $\mathcal{N}_0^d$ is non-negative and is written out only in terms of the eigenvalues of the covariance matrix.

The organization of the paper is as follows. We start with a review of the L^2-Wasserstein geometry and the Riemannian length space in Section 2. Then we prove Theorem 1.1 in Section 3, using the approximate expression of sectional curvature. We demonstrate the correspondence between our results and previously obtained result in Section 4.

Acknowledgements

The author would like to express her gratitude to Professor Sumio Yamada, for his advice, support and encouragement. His valuable suggestions were essential to the completion of this paper.

2. *Preliminaries*

2.1. L^2-*Wasserstein space*

We first review L^2-Wasserstein spaces (see [**16**].) Given a complete metric space (X, d), we denote by $\mathcal{P}_2(X)$ the set of probability measures with finite second moments on X.

Definition 2.1. *For $\mu,\ \nu \in \mathcal{P}_2(X)$, a transport plan π between μ and ν is a Borel probability measure on $X \times X$ with marginals μ and ν, that is,*

$$\pi[A \times X] = \mu[A], \quad \pi[X \times A] = \nu[A] \quad \text{for all Borel sets } A \text{ in } X.$$

Let $\Pi(\mu,\nu)$ be the set of transport plans between μ and ν, then the L^2-Wasserstein distance between μ and ν is defined by

$$W_2(\mu,\nu)^2 = \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d(x,y)^2 d\pi(x,y).$$

The L^2-Wasserstein distance actually becomes a distance. We call the pair $(\mathcal{P}_2(X), W_2)$ the L^2-Wasserstein space over X. A transport plan which achieves the infimum is called optimal. Optimal transport plans on Euclidean spaces are characterized by the following properties.

Theorem 2.2 ([**2**],[**6**]). *Let μ and ν be Borel probability measures on $\mathbb{R}^d$. If μ is absolutely continuous with respect to Lebesgue measure, then*

(i) *there exists a convex function ψ on $\mathbb{R}^d$ whose gradient $\nabla\psi$ pushes μ forward to ν.*
(ii) *this gradient is uniquely determined (μ-almost everywhere.)*
(iii) *the joint measure $\pi = (\mathrm{id} \times \nabla\psi)_\sharp \mu$ is optimal.*
(iv) *π is the only optimal measure in $\Pi(\mu,\nu)$ unless $W_2(\mu,\nu) = +\infty$.*

Here the push forward measure of μ through measurable map $f : \mathbb{R}^d \to \mathbb{R}^d$, denoted by $f_\sharp\mu$, is defined by $f_\sharp\mu[A] = \mu[f^{-1}(A)]$ for all Borel sets A in $\mathbb{R}^d$.

McCann [**9**] obtained the optimal transport plans between Gaussian measures on $\mathbb{R}^d$ and showed that the displacement interpolation between any two Gaussian measures is also a Gaussian measure. Namely, $\mathcal{N}^d$ is a geodesically convex subset of the L^2-Wasserstein space.

Lemma 2.3 ([**9**, Example 1.7]). *For $X \in \mathrm{Sym}^+(d,\mathbb{R})$, we define a symmetric positive definite matrix $X^{1/2}$ so that $X^{1/2} \cdot X^{1/2} = X$. For $N(m,V)$ and $N(n,U)$, define a symmetric positive definite matrix*

$$W = (w_{ij}) = U^{\frac{1}{2}}(U^{\frac{1}{2}}VU^{\frac{1}{2}})^{-\frac{1}{2}}U^{\frac{1}{2}}$$

and the related function

$$\mathcal{W}(x) = \frac{1}{2}\langle x-m, W(x-m)\rangle + \langle x, n\rangle.$$

We denote the gradient of $\mathcal{W}$ by $\nabla\mathcal{W}$. Then, $(\mathrm{id}, \nabla\mathcal{W})_\sharp N(m,V)$ is the optimal transport between $N(m,V)$ and $N(n,U)$. If we moreover set

$$l(t) = (1-t)m + tn, \quad W(t) = ((1-t)E + tW)V((1-t)E + tW),$$

then $\{N(l(t), W(t))\}_{t\in[0,1]}$ is a geodesic from $N(m,V)$ to $N(n,U)$.

Lemma 2.3 enables us to obtain the L^2-Wasserstein distance on $\mathcal{N}_0^d$.

Theorem 2.4 ([**4**], [**7**], [**10**], [**11**]). *For $N(m,V)$ and $N(n,U)$, we get*

$$W_2(N(m,V),N(n,U))^2 = |m-n|^2 + \mathrm{tr}V + \mathrm{tr}U - 2\mathrm{tr}\left(U^{\frac{1}{2}}VU^{\frac{1}{2}}\right)^{\frac{1}{2}}.$$

We call $\mathcal{W}$ above a (unique) linear transform between $N(m,V)$ and $N(n,U)$. Let $O(d)$ be the set of orthogonal matrices of size d. For $P \in O(d)$, we denote by $\mathcal{N}^d(P)$ the subset of $\mathcal{N}^d$ whose covariance matrices are diagonalized by P.

Corollary 2.5. *For any $P \in O(d)$, $(\mathcal{N}^d(P), W_2)$ is isometric to $\mathbb{R}^d \times (\mathbb{R}_{>0})^d$.*

Proof. For $N(m,V), N(n,U) \in \mathcal{N}^d(P)$, there uniquely exist $\{\lambda_i\}_{i=1}^d, \{\sigma_i\}_{i=1}^d \subset \mathbb{R}_{>0}$ such that

$$V = P\mathrm{diag}[\lambda_1,\dots,\lambda_d]^{\mathrm{T}}P, \quad U = P\mathrm{diag}[\sigma_1,\dots,\sigma_d]^{\mathrm{T}}P.$$

By Theorem 2.4, we have

$$W_2(N(m,V),N(n,U))^2 = |m-n|^2 + \sum_{i=1}^d (\lambda_i - \sigma_i)^2.$$

Therefore a map identifying $N(m,V)$ with $(m,(\lambda_1,\dots,\lambda_d))$ is an isometry from $\mathcal{N}^d(P)$ to $\mathbb{R}^d \times (\mathbb{R}_{>0})^d$. □

Remark 2.6. *In the L^2-Wasserstein geometry, $\mathcal{N}^1$ is isometric to a Euclidean upper half plane. While in the Fisher geometry, $\mathcal{N}^1$ is isometric to a hyperbolic plane with constant sectional curvature $-1/2$* (see [**1**].)

2.2. *Riemannian length space*

Next, we give $\mathcal{N}^d$ an L^2-Wasserstein metric. See [**3**] for more detail.

Definition 2.7. *Let $\langle\cdot,\cdot\rangle_y$ and $|\cdot|_y$ denote an inner product and a norm on a vector space $\mathcal{H}_y$. A subset M of a length space (N, dist) is called Riemannian if each $x \in M$ is associated with a map $\epsilon\mathrm{xp}_x : \mathcal{H}_x \to N$ defined on some inner product space $\mathcal{H}_x$ which gives a surjection from a star-shaped subset $\mathcal{K}_x \subset \mathcal{H}_x$ onto M such that the curve $x_s = \epsilon\mathrm{xp}_x(sp)$ defines an (affinely parameterized) minimizing geodesic $[0,1] \ni s \mapsto x_s$ linking $x = x_0$ to $y = x_1$ for each $p \in \mathcal{K}_x$. We moreover assume that there exists $q \in \mathcal{K}_y$ such that $x_s = \epsilon\mathrm{xp}_y(1-s)q$ and*

$$\mathrm{dist}(\epsilon\mathrm{xp}_x u, \epsilon\mathrm{xp}_y v)^2 \le \mathrm{dist}(x,y)^2 - 2\langle v,q\rangle_y - 2\langle u,p\rangle_x + o(\sqrt{|u|_x^2 + |v|_y^2}),$$

for all $u \in \mathcal{H}_x$ and $v \in \mathcal{H}_y$ as $|u|_x + |v|_y \to 0$. Dependence of these structures on the base points x and y may be suppressed when it can be inferred from the context.

It was shown in [**3**, Proposition 4.1] that $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ forms a Riemannian length space with the following methods.

Take $(N, \mathrm{dist}) = (\mathcal{P}_2(\mathbb{R}^d), W_2)$ as our complete length space and the subset $M = \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$. Fix $\rho \in M$. Let $\mathrm{spt}(\rho)$ denote smallest closed subset of $\mathbb{R}^d$ containing the full mass of ρ, and let $\Omega_\rho \subset \mathbb{R}^d$ denote the interior of the convex hull of $\mathrm{spt}(\rho)$. We take $\mathcal{H}_\rho = \mathcal{H}^{1,2}(\mathbb{R}^d, d\rho) \subset C^{0,1}_{\mathrm{loc}}(\Omega_\rho)$ to consist of those locally Lipschitz continuous functions on Ω_ρ whose first derivatives lie in

the weighted space $L^2(\mathbb{R}^d, d\rho; \mathbb{R}^d)$, modulo equivalence with respect to semi-norm

$$\langle \psi, \psi \rangle_\rho = \int_{\Omega_\rho} |\nabla \psi(x)|^2 d\rho(x).$$

And the exponential map is defined by

$$\epsilon\mathrm{exp}_\rho s\psi = [\mathrm{id} + s\nabla\psi]_\sharp \rho.$$

Furthermore, they remarked if $M' \subset M$ is geodesically convex, meaning any geodesic lies in M' whenever its endpoints do, then M' is itself a Riemannian length space with the same tangent space and the exponential map as those of M, but the star-shaped subset is given by

$$\mathcal{K}'_x = \{p \in \mathcal{K}_x \bigm| \exp_x p \in M'\}.$$

Lemma 2.3 implies that $\mathcal{N}_0^d$ is a geodesically convex subset of $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$. Therefore $\mathcal{N}_0^d$ should also be a Riemannian length space, especially a Riemannian manifold. If $\mathcal{W}$ is the linear transform between $N(V)$ and $N(U)$, then $\epsilon\mathrm{exp}_{N(V)} s\psi = N(U)$ if and only if $\nabla\psi$ is same as $\nabla(\mathcal{W} - |\cdot|^2/2)$. Identifying linear transforms as their coefficients, we treat the tangent vector space at each point as $\mathrm{Sym}(d, \mathbb{R})$. Namely, we can identify the tangent space at $N(V)$ to $\mathcal{N}_0^d$ with $\mathrm{Sym}(d, \mathbb{R})$ based on the idea

$$\exp_{N(V)} tX = N(U(t)), \text{where } U(t) = ((1-t)E + tX)V((1-t)E + tX).$$

Moreover, the inner product becomes a Riemannian metric g on $\mathcal{N}_0^d$, whose Riemannian distance coincides with the L^2-Wasserstein distance. Its expression of g is given by

$$g_{N(V)}(tX, tX) = \int_{\mathbb{R}^d} |tXx|^2 dN(V)(x) = t^2 \mathrm{tr} XVX.$$

Theorem 2.8. *Let $\mathcal{N}_0^d$ be the space of Gaussian measures with mean 0 over $\mathbb{R}^d$. Then, $\mathcal{N}_0^d$ becomes a C^∞-Riemannian manifold of dimension $d(d+1)/2$ and there exists the L^2-Wasserstein metric g. If we identify the tangent space at $N(V)$ to $\mathcal{N}_0^d$ with $\mathrm{Sym}(d, \mathbb{R})$ by*

$$\exp_{N(V)} tX = N(U(t)), \textit{where } U(t) = ((1-t)E + tX)V((1-t)E + tX),$$

then it is explicitly given by

$$g_{N(V)}(X, Y) = \mathrm{tr} XVY.$$

Theorem 2.8 shows that $\{e_+, e_{ij}, f_{ij}\}_{1 \le i < j \le d}$ is a set of normal vectors.

3. *Proof of Theorem 1.1*

In order to calculate the sectional curvatures of $(\mathcal{N}_0^d, g)$, where g is the L^2-Wasserstein metric, we need some lemmas.

Lemma 3.1 ([**5**, Theorem 3.68]). *Let (M, g) be a Riemannian manifold. For any $p \in M$, $\{u, v\}$ is an orthonormal basis of a 2-plane in the tangent space at p. Let*

$$C_r(\theta) = \exp_p r(u\cos\theta + v\sin\theta),$$

and $L(r)$ be the length of the curve C_r. Then the function $L(r)$ admits an asymptotic expansion

$$L(r) = 2\pi r\left(1 - \frac{K(u,v)}{6} r^2 + o(r^2)\right), \quad \textit{as} \quad r \searrow 0,$$

where $K(u, v)$ is the sectional curvature of the 2-plane spanned by $\{u, v\}$.

LEMMA 3.2. *For $A, B \in \{e_+, e_{ij}, f_{ij}\}_{1\le i<j\le d}$, $0 \le r \ll 1$ and $\theta \in [0, 2\pi]$,*

$$C_r(\theta) = \exp_{N(V)} r(\cos\theta \cdot A + \sin\theta \cdot B)$$

is a Gaussian measure whose covariance matrix $X = X(r,\theta) = (x_{\alpha\beta})$ is given by

$$X = [E + r(\cos\theta \cdot A + \sin\theta \cdot B)] \cdot V \cdot [E + r(\cos\theta \cdot A + \sin\theta \cdot B)], \tag{3.1}$$

where E is the identity matrix.

Proof. It is clear by Lemma 2.3. □

Proof of (1) and (5)
If we choose $A = e_+, B = e_{ij}$ or $A = e_{ij}, B = e_{kl}$ in (3.1), then $N(X)$ belongs to $\mathcal{N}_0^d(P)$. Since $\mathcal{N}_0^d(P)$ is a flat manifold by Corollary 2.5, the curvatures vanish.

A strategy for proving the remaining case is as follows. We first calculate

$$W(\theta_0,\theta) = W_2(C_r(\theta_0), C_r(\theta))^2 \quad \text{and} \quad W(\theta_0) = \lim_{\theta\to\theta_0} \frac{W(\theta_0,\theta)}{\theta^2}.$$

Then we get

$$L(r) = \int_0^{2\pi} W(\theta)^{\frac{1}{2}} d\theta.$$

Finally we use Lemma 3.1 to obtain the expression of the sectional curvatures. Without loss of generality, we may assume $P = E$. That is to say,

$$e_+ = \frac{E_{11} + E_{dd}}{\sqrt{\lambda_1 + \lambda_d}},\ e_{ij} = \frac{E_{ii} - E_{jj}}{\sqrt{\lambda_i + \lambda_j}},\ f_{ij} = \frac{E_{ij} + E_{ji}}{\sqrt{\lambda_i + \lambda_j}},\ \text{and } V = \mathrm{diag}[\lambda_1, \ldots, \lambda_d],$$

because we have

$$W(\theta_0,\theta) = \mathrm{tr}X(r,\theta_0) + \mathrm{tr}X(r,\theta) - 2\mathrm{tr}\left(X(r,\theta_0)^{\frac{1}{2}} X(r,\theta) X(r,\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}}$$

and the value is invariant under taking conjugation with any orthogonal matrix P.

For a general symmetric positive definite matrix X, it is hard to get a concrete expression of $X^{1/2}$. But if the matrix is size of 2×2, the next lemma enables us to obtain the value of the trace and the determinant of $X^{1/2}$.

LEMMA 3.3. *Let $M \in \mathrm{Sym}(2, \mathbb{R})$, then*

$$(\mathrm{tr}M)^2 = \mathrm{tr}M^2 + 2\det M.$$

Proof. Setting

$$M = \begin{pmatrix} a & c \\ c & b \end{pmatrix},\ \text{we obtain } M^2 = \begin{pmatrix} a^2 + c^2 & c(a+b) \\ c(a+b) & b^2 + c^2 \end{pmatrix}.$$

Therefore, we get

$$\mathrm{tr}M^2 + 2\det M = a^2 + b^2 + 2c^2 - 2(ab - c^2) = (a+b)^2 = (\mathrm{tr}M)^2.$$

□

We set

$$c_{ij}(r,\theta) = \frac{r\cos\theta}{\sqrt{\lambda_i + \lambda_j}},\ s_{ij}(r,\theta) = \frac{r\sin\theta}{\sqrt{\lambda_i + \lambda_j}}$$

for $1 \le i, j \le d$, $\theta \in [0, 2\pi]$ and sufficiently small $r \ge 0$.

Proof of (2) and (8)
For (2), we take $A = f_{1d}$, $B = e_+$ and $I = \{1, d\}$, whereas, for (8), take $A = f_{ij}$, $B = e_{ij}$ and $I = \{i, j\}$. Then we notice that for any $\alpha, \beta \notin I$, (α, β)-components of X are independent of the variables r and θ. If we set

$$\widetilde{X}(\theta) = \begin{pmatrix} x_{\alpha\alpha} & x_{\alpha\beta} \\ x_{\beta\alpha} & x_{\beta\beta} \end{pmatrix}$$

for $\{\alpha, \beta\} = I$, we obtain

$$W(\theta_0, \theta) = \mathrm{tr}\widetilde{X}(\theta_0) + \mathrm{tr}\widetilde{X}(\theta) - 2\mathrm{tr}\left(\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}}. \tag{3.2}$$

For (2), using Lemma 3.3, we conclude

$$W(\theta_0, \theta) = 4r^2 \sin^2(\theta - \theta_0) \quad \text{and} \quad \lim_{\theta \to \theta_0} \frac{W(\theta_0, \theta)}{(\theta - \theta_0)^2} = r^2.$$

It follows that $L(r) = 2\pi r$, proving $K(e_+, f_{1d}) = 0$.

For (8), in a similar way, we have

$$W(\theta_0, \theta) = 4r^2 \sin^2 \frac{1}{2}(\theta - \theta_0) - \frac{4r^4 \lambda_i \lambda_j \sin^2(\theta - \theta_0)}{(\lambda_i + \lambda_j)^2 a_r(\theta_0, \theta)} + o(|\theta - \theta_0|^2),$$

where

$$\begin{aligned} a_r(\theta_0, \theta) =& \lambda_i \left[(1 + c_{ij}(r, \theta_0))(1 + c_{ij}(r, \theta)) + s_{ij}(r, \theta_0) s_{ij}(r, \theta)\right] \\ &+ \lambda_j \left[(1 - c_{ij}(r, \theta_0))(1 - c_{ij}(r, \theta)) + s_{ij}(r, \theta_0) s_{ij}(r, \theta)\right]. \end{aligned}$$

Since the limit of $a_r(\theta_0, \theta)$ exists as $\theta \to \theta_0$ and

$$a_r(\theta_0, \theta_0) = (\lambda_i + \lambda_j)(1 + r^2) + 2(\lambda_i - \lambda_j) r \cos\theta_0,$$

we have

$$\lim_{\theta \to \theta_0} \frac{W(\theta_0, \theta)}{(\theta - \theta_0)^2} = r^2 - \frac{4r^4 \lambda_i \lambda_j}{(\lambda_i + \lambda_j)^2 a_r(\theta_0, \theta_0)}.$$

It follows that

$$\begin{aligned} L(r) &= \int_0^{2\pi} r \left(1 - \frac{4r^2 \lambda_i \lambda_j}{(\lambda_i + \lambda_j)^2 a_r(\theta, \theta)}\right)^{\frac{1}{2}} d\theta \\ &= \int_0^{2\pi} r \left(1 - \frac{1}{2} \frac{4r^2 \lambda_i \lambda_j}{(\lambda_i + \lambda_j)^2 a_r(\theta, \theta)} + o(r^2)\right) d\theta. \end{aligned}$$

Because $a_0(\theta, \theta) = \lambda_i + \lambda_j$, using Lemma 3.1 and the bounded convergence theorem, we obtain

$$K(e_{ij}, f_{ij}) = \frac{12\lambda_i \lambda_j}{(\lambda_i + \lambda_j)^3}.$$

Proof of (3) and (7)
For (3), assuming $i = 1$, take $A = e_+$, $B = f_{1j}$ and $I = \{1, j, d\}$, whereas, for (7), assuming $j < k$, take $A = e_{ik}$, $B = f_{ij}$ and $I = \{i, j, k\}$. Since for any $\alpha, \beta \notin I$, (α, β)-components of X are independent of the variables r and θ, we obtain

$$\begin{aligned} &W(\theta_0, \theta) \\ &= \mathrm{tr}\widetilde{X}(\theta_0) + \mathrm{tr}\widetilde{X}(\theta) - 2\mathrm{tr}\left(\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}} \\ &= \mathrm{tr}\widetilde{Y}(\theta_0) + \mathrm{tr}\widetilde{Y}(\theta) - 2\mathrm{tr}\left(\widetilde{Y}(\theta_0)^{\frac{1}{2}}\widetilde{Y}(\theta)\widetilde{Y}(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}} + \frac{r^2 \lambda_\gamma}{\lambda_\alpha + \lambda_\gamma}(\cos\theta - \cos\theta_0)^2, \end{aligned}$$

where

$$\widetilde{X}(\theta) = \begin{pmatrix} x_{\alpha\alpha} & x_{\alpha\beta} & x_{\alpha\gamma} \\ x_{\beta\alpha} & x_{\beta\beta} & x_{\beta\gamma} \\ x_{\gamma\alpha} & x_{\gamma\beta} & x_{\gamma\gamma} \end{pmatrix} = \begin{pmatrix} \widetilde{Y}(\theta) & {}^{\mathrm{T}}\mathbf{0} \\ \mathbf{0} & \lambda_\gamma(1+c_{\alpha\gamma}(r,\theta))^2 \end{pmatrix}, \quad \mathbf{0} = (0,0)$$

and $\{\alpha,\beta,\gamma\} = I$. Using Lemma 3.3, we conclude

$$W(\theta_0,\theta) = 4r^2 \sin^2 \frac{1}{2}(\theta-\theta_0) - \frac{r^4}{a_r(\theta,\theta_0)} \frac{\lambda_\alpha\lambda_\beta \sin^2(\theta-\theta_0)}{(\lambda_\alpha+\lambda_\beta)(\lambda_\alpha+\lambda_\gamma)} + o(\theta^2),$$

where

$$a_r(\theta,\theta_0) = \lambda_\alpha(1+c_{\alpha\gamma}(r,\theta_0))(1+c_{\alpha\gamma}(r,\theta)) + r^2 \sin\theta_0 \sin\theta + \lambda_\beta.$$

Since the limit of $a_r(\theta_0,\theta)$ exists as $\theta \to \theta_0$, we have

$$\lim_{\theta\to\theta_0} \frac{W(\theta_0,\theta)}{(\theta-\theta_0)^2} = r^2 - \frac{r^4}{a_r(\theta_0,\theta_0)} \frac{\lambda_\alpha\lambda_\beta}{(\lambda_\alpha+\lambda_\beta)(\lambda_\alpha+\lambda_\gamma)}.$$

It follows that

$$L(r) = \int_0^{2\pi} r \left(1 - \frac{1}{2} \frac{r^2}{a_r(\theta,\theta)} \frac{\lambda_\alpha\lambda_\beta}{(\lambda_\alpha+\lambda_\beta)(\lambda_\alpha+\lambda_\gamma)} + o(r^2)\right) d\theta.$$

Because $a_0(\theta,\theta) = (\lambda_\alpha+\lambda_\beta)$, using Lemma 3.1 and the bounded convergence theorem, we obtain

$$K(A,B) = \frac{3\lambda_\alpha\lambda_\beta}{(\lambda_\alpha+\lambda_\beta)^2(\lambda_\alpha+\lambda_\gamma)}.$$

We can prove the case of $i \neq 1$ and $j = d$ in a similar way.

Proof of (4),(6) and (9)

We take (A,B) in (3.1) as (e_+, f_{kl}) $(\{1,d\} \cup \{k,l\} = \emptyset)$, (e_{ij}, f_{kl}) $(\{i,j\} \cup \{k,l\} = \emptyset)$ and (f_{ij}, f_{kl}) $(\{i,j\} \cup \{k,l\} = \emptyset)$ in this order. Moreover we set $I = \{1,d\}$ in the case (4) and $I = \{i,j\}$ in the case of (6) and (9). We notice that for any $\alpha,\beta \notin I$, (α,β)-components of X are independent of the variables r and θ. If we set

$$\widetilde{X}_c(\theta) = \begin{pmatrix} x_{\alpha\alpha} & x_{\alpha\beta} \\ x_{\beta\alpha} & x_{\beta\beta} \end{pmatrix}, \ \widetilde{X}_s(\theta) = \begin{pmatrix} x_{kk} & x_{kl} \\ x_{lk} & x_{ll} \end{pmatrix},$$

we obtain

$$\begin{aligned} W(\theta_0,\theta) =& \mathrm{tr}\widetilde{X}_c(\theta_0) + \mathrm{tr}\widetilde{X}_c(\theta) - 2\mathrm{tr}\left(\widetilde{X}_c(\theta_0)^{\frac{1}{2}}\widetilde{X}_c(\theta)\widetilde{X}_c(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}} \\ &+ \mathrm{tr}\widetilde{X}_s(\theta_0) + \mathrm{tr}\widetilde{X}_s(\theta) - 2\mathrm{tr}\left(\widetilde{X}_s(\theta_0)^{\frac{1}{2}}\widetilde{X}_s(\theta)\widetilde{X}_s(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}}, \end{aligned} \tag{3.3}$$

where $\{\alpha,\beta\} = I$. Using Lemma 3.3, we conclude

$$\lim_{\theta\to\theta_0} \frac{W(\theta_0,\theta)}{(\theta-\theta_0)^2} = \left(\lim_{\theta\to\theta_0} \frac{r\sin(\theta-\theta_0)}{\theta-\theta_0}\right)^2 = r^2.$$

It follows that $L(r) = 2\pi r$ and $K(A,B) = 0$.

Proof of (10)

Without loss of generality, we may assume $j < k$. Taking A and B as f_{ij} and f_{ik} in (3.1) respectively. We notice that for any $\alpha,\beta \notin \{i,j,k\}$, (α,β)-components of X are independent

of the variables r and θ. If we set

$$\widetilde{X}(\theta) = \begin{pmatrix} x_{ii} & x_{ij} & x_{ik} \\ x_{ji} & x_{jj} & x_{jk} \\ x_{ki} & x_{kj} & x_{kk} \end{pmatrix}$$
$$= \begin{pmatrix} \lambda_i + \lambda_j c_{ij}(r,\theta)^2 + \lambda_k s_{ik}(r,\theta)^2 & (\lambda_i+\lambda_j)c_{ij}(r,\theta) & (\lambda_i+\lambda_k)s_{ik}(r,\theta) \\ (\lambda_i+\lambda_j)c_{ij}(r,\theta) & \lambda_j + \lambda_i c_{ij}(r,\theta)^2 & \lambda_i c_{ij}(r,\theta)s_{ik}(r,\theta) \\ (\lambda_i+\lambda_k)s_{ik}(r,\theta) & \lambda_i c_{ij}(r,\theta)s_{ik}(r,\theta) & \lambda_k + \lambda_i s_{ik}(r,\theta)^2 \end{pmatrix},$$

we obtain

$$W(\theta_0,\theta) = \mathrm{tr}\widetilde{X}(\theta_0) + \mathrm{tr}\widetilde{X}(\theta) - 2\mathrm{tr}\left(\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}}. \tag{3.4}$$

For the value of the last term in (3.4), Lemma 3.3 can not be used as the size of matrices is 3×3.

We define some notations:

$$\begin{aligned}
&A = A_{\theta_0}(\theta) = \widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}} \\
&B = B_{\theta_0}(\theta) = \left(\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}}\right)^{\frac{1}{2}} \\
&\{\sigma_\alpha = \sigma_{\theta_0}(\theta)_\alpha\}_{\alpha=1}^3 : \text{eigenvalues of } B \\
&f_{\theta_0}(\theta) = \mathrm{tr}B = \sigma_1+\sigma_2+\sigma_3 \\
&g_{\theta_0}(\theta) = \mathrm{tr}A = \sigma_1^2+\sigma_2^2+\sigma_3^2 \\
&h_{\theta_0}(\theta) = \sigma_1\sigma_2+\sigma_2\sigma_3+\sigma_3\sigma_1 \\
&\varphi_{\theta_0}(\theta) = \sigma_1^2\sigma_2^2+\sigma_2^2\sigma_3^2+\sigma_3^2\sigma_1^2 \\
&D_{\theta_0}(\theta) = \det B = (\det A)^{\frac{1}{2}} = \sigma_1\sigma_2\sigma_3
\end{aligned}$$

Rewriting (3.4) with the Taylor approximation of $f_{\theta_0}(\cdot)$ at θ_0, we obtain

$$W(\theta_0,\theta) = -2f'_{\theta_0}(\theta_0)(\theta-\theta_0) - f''_{\theta_0}(\theta_0)(\theta-\theta_0)^2 + o(|\theta-\theta_0|^2).$$

Since we can get the values of g, φ and D without information of $X^{1/2}$, we compute f' and f'' by using these values.

We calculate $f'_{\theta_0}(\theta_0)$ first. Differentiating $B_{\theta_0}(\theta)\cdot B_{\theta_0}(\theta) = A_{\theta_0}(\theta)$ with respect to θ, we have

$$B'_{\theta_0}(\theta)B_{\theta_0}(\theta) + B_{\theta_0}(\theta)B'_{\theta_0}(\theta) = A'_{\theta_0}(\theta) = \widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}'(\theta)\widetilde{X}(\theta_0)^{\frac{1}{2}}.$$

After multiplying $B_{\theta_0}(\theta)^{-1}$ from the left, taking the trace gives

$$\mathrm{tr}B'_{\theta_0}(\theta_0) + \mathrm{tr}(B_{\theta_0}(\theta_0)B'_{\theta_0}(\theta_0)B_{\theta_0}(\theta_0)^{-1}) = 2f'_{\theta_0}(\theta_0)$$

at $\theta=\theta_0$. Because $\mathrm{tr}\widetilde{X}(\theta)$ is constant, at $\theta=\theta_0$ the right hand side is equal to

$$\mathrm{tr}(\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}'(\theta_0)\widetilde{X}(\theta_0)^{\frac{1}{2}}\widetilde{X}(\theta_0)^{-1}) = \mathrm{tr}\widetilde{X}'(\theta_0) = \left(\mathrm{tr}\widetilde{X}(\theta)\right)'\Big|_{\theta=\theta_0} = 0.$$

Therefore we conclude

$$f'_{\theta_0}(\theta_0) = 0. \tag{3.5}$$

Next we compute $f''_{\theta_0}(\theta_0)$. Differentiating $f^2 = g + 2h$ at $\theta=\theta_0$, we have

$$2f_{\theta_0}(\theta_0)f'_{\theta_0}(\theta_0) = g'_{\theta_0}(\theta_0) + 2h'_{\theta_0}(\theta_0),$$

proving

$$2h'_{\theta_0}(\theta_0) = -g'_{\theta_0}(\theta_0)$$

because of (3.5). Differentiating once more,

$$f''_{\theta_0}(\theta) = -\frac{f'_{\theta_0}(\theta)}{2f_{\theta_0}(\theta)^2}\left(g'_{\theta_0}(\theta) + 2h'_{\theta_0}(\theta)\right) + \frac{g''_{\theta_0}(\theta) + 2h''_{\theta_0}(\theta)}{2f_{\theta_0}(\theta)}.$$

Because of (3.5), we get at $\theta = \theta_0$

$$f''_{\theta_0}(\theta_0) = \frac{g''_{\theta_0}(\theta_0) + 2h''_{\theta_0}(\theta_0)}{2f_{\theta_0}(\theta_0)}. \tag{3.6}$$

We compute directly

$$g_{\theta_0}(\theta) = \sum_{\alpha,\beta\in\{i,j,k\}} x_{\alpha\beta}(\theta_0)x_{\beta\alpha}(\theta).$$

This enables us to get the derivatives of $g_{\theta_0}(\theta)$. Because $B_{\theta_0}(\theta_0) = X(\theta_0)$, using the relation

$$\det(tE - B) = t^3 - t^2\cdot f + t\cdot h - D,$$

we have

$$h_{\theta_0}(\theta_0) = \sum_{\substack{\alpha,\beta\in\{i,j,k\}\\ \alpha\neq\beta}} \left(x_{\alpha\alpha}(\theta_0)x_{\beta\beta}(\theta_0) - x_{\alpha\beta}(\theta_0)^2\right).$$

While it is hard to compute $B_{\theta_0}(\theta)$ directly, it is also hard to know the value of $h_{\theta_0}(\theta)$. We want to derive $h''_{\theta_0}(\theta)$ without the information of $B_{\theta_0}(\theta)$. So differentiating $h^2 = \varphi + 2Df$ twice, we have

$$2(h'_{\theta_0}(\theta))^2 + 2h_{\theta_0}(\theta)h''_{\theta_0}(\theta) = \varphi''_{\theta_0}(\theta) + 4D'_{\theta_0}(\theta)f'_{\theta_0}(\theta) + 2D''_{\theta_0}(\theta)f_{\theta_0}(\theta) + 2D_{\theta_0}(\theta)f''_{\theta_0}(\theta).$$

At $\theta = \theta_0$, we have

$$h''_{\theta_0}(\theta_0) = -\frac{g'_{\theta_0}(\theta_0)^2}{4h_{\theta_0}(\theta_0)} + \frac{\varphi''_{\theta_0}(\theta_0) + 2D''_{\theta_0}(\theta_0)f_{\theta_0}(\theta_0) + 2D_{\theta_0}(\theta_0)f''_{\theta_0}(\theta_0)}{2h_{\theta_0}(\theta_0)}. \tag{3.7}$$

In order to analyze (3.7), we consider $D_{\theta_0}(\theta)$ and $\varphi_{\theta_0}(\theta)$. From the definition, we can compute $D_{\theta_0}(\theta)$ directly as

$$D_{\theta_0}(\theta) = \lambda_i\lambda_j\lambda_k\left[1 - (c_{ij}(r,\theta_0)^2 + s_{ik}(r,\theta_0)^2)\right]\left[1 - (c_{ij}(r,\theta)^2 + s_{ik}(r,\theta)^2)\right].$$

We next consider $\varphi_{\theta_0}(\theta)$. Using the equation

$$\det(tE - A) = \det\widetilde{X}(\theta_0)\cdot\det(t\widetilde{X}(\theta_0)^{-1} - \widetilde{X}(\theta)),$$

and the relation

$$\det(tE - A) = t^3 - t^2\cdot g(\theta) + t\cdot\varphi - D^2,$$

we conclude

$$\varphi_{\theta_0}(\theta) = \det(\widetilde{X}(\theta_0)\widetilde{X}(\theta))\cdot\mathrm{tr}(Y(\theta_0)Y(\theta)),\ \text{where } Y(\theta) = \widetilde{X}(\theta)^{-1}. \tag{3.8}$$

Since (3.8) depends only on $\widetilde{X}(\theta)$, we can obtain the value of $\varphi_{\theta_0}(\theta)$. Therefore we can now specify the value of $h''_{\theta_0}(\theta_0)$ in(3.7).

Inserting (3.7) into (3.6), we obtain

$$W(\theta_0,\theta) = -f''_{\theta_0}(\theta_0) + o(|\theta-\theta_0|^2) = -\frac{\beta_r(\theta_0)}{\alpha_r(\theta_0)} + o(|\theta-\theta_0|^2),$$

where

$$\begin{aligned}\alpha_r(\theta_0) =& 2\left[f_{\theta_0}(\theta_0)h_{\theta_0}(\theta_0) - D_{\theta_0}(\theta_0)\right] \\ =& 2(\lambda_i+\lambda_j)(\lambda_j+\lambda_k)(\lambda_k+\lambda_i) \\ &+ r^2\left[\lambda_j^2+\lambda_k^2+4\lambda_j\lambda_k+\lambda_i\lambda_j+\lambda_i\lambda_k+(\lambda_j-\lambda_k)(\lambda_j+\lambda_k+3\lambda_i)\cos\theta_0\right],\end{aligned} \tag{3.9}$$

$$\begin{aligned}\beta_r(\theta_0) =& h_{\theta_0}(\theta_0)g''_{\theta_0}(\theta_0) - \frac{1}{2}g'_{\theta_0}(\theta_0)^2 + \varphi''_{\theta_0}(\theta_0) + 2D''_{\theta_0}(\theta_0) \\ =& -2r^2(\lambda_i+\lambda_j)(\lambda_j+\lambda_k)(\lambda_k+\lambda_i) \\ &- r^4\left[(\lambda_j+\lambda_k)(\lambda_i+\lambda_j+\lambda_k)+(\lambda_j-\lambda_k)(\lambda_j+\lambda_k+3\lambda_i)\cos\theta_0\right].\end{aligned} \tag{3.10}$$

Therefore we have

$$W(\theta_0) = \lim_{\theta\to\theta_0}\frac{W(\theta,\theta_0)}{\theta^2} = -\frac{\beta_r(\theta_0)}{\alpha_r(\theta_0)}.$$

If we set

$$\begin{aligned}L &= 2(\lambda_i+\lambda_j)(\lambda_j+\lambda_k)(\lambda_k+\lambda_i), \\ a &= \lambda_j^2+\lambda_k^2+4\lambda_j\lambda_k+\lambda_i\lambda_j+\lambda_i\lambda_k+(\lambda_j-\lambda_k)(\lambda_j+\lambda_k+3\lambda_i)\cos\theta, \\ b &= (\lambda_j+\lambda_k)(\lambda_i+\lambda_j+\lambda_k)+(\lambda_j-\lambda_k)(\lambda_j+\lambda_k+3\lambda_i)\cos\theta,\end{aligned}$$

we have

$$L(r) = \int_0^{2\pi} r\left(1+\frac{r^2(b-a)}{2(L+r^2a)}+o(r^2)\right)d\theta.$$

Using Lemma 3.1 and the bounded convergence theorem, we obtain

$$2\pi\frac{K(u,v)}{6} = \int_0^{2\pi}\lim_{r\searrow 0}\frac{a-b}{2(L+r^2a)}d\theta = 2\pi\frac{a-b}{2L},$$

which implies that

$$K(f_{ij},f_{ik}) = \frac{3\lambda_k\lambda_j}{(\lambda_i+\lambda_j)(\lambda_j+\lambda_k)(\lambda_k+\lambda_i)}.$$

This completes the proof of Theorem 1.1.

4. *Remarks to Theorem 1.1*

In this section we consider the case $d = 2$ in particular.

4.1. *Geometric interpolations of Theorem 1.1*

LEMMA 4.1. *Any $V = (v_{ij}) \in \mathrm{Sym}(2,\mathbb{R})$ is diagonalized by some special orthogonal matrix. In other word, there exists some $\theta \in \mathbb{R}$ such that the rotation matrix*

$$R(\theta) = \begin{pmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix}$$

diagonalizes V.

Proof. In the case of $v_{12} = 0$, we set $\theta = 0$. While in the case of $v_{12} \neq 0$, since

$${}^{\mathrm{T}}R(\theta)VR(\theta) = \begin{pmatrix} v_{11}\cos^2\theta + v_{12}\sin 2\theta + v_{22}\sin^2\theta & v_{12}\cos 2\theta + 2^{-1}(v_{11}-v_{22})\sin 2\theta \\ v_{12}\cos 2\theta + 2^{-1}(v_{11}-v_{22})\sin 2\theta & v_{22}\cos^2\theta - v_{12}\sin 2\theta + v_{11}\sin^2\theta\end{pmatrix},$$

${}^{\mathrm{T}}R(\theta)VR(\theta)$ is a diagonal matrix if and only if

$$v_{12}\cos 2\theta = -2^{-1}(v_{11}-v_{22})\sin 2\theta \Leftrightarrow \cot 2\theta = (v_{11}-v_{22})/2v_{12}. \tag{4.1}$$

Because $\cot 2\theta$ can take any value, (4.1) always holds true. $\square$

For $\alpha, \beta > 0$, we denote

$$(\alpha, \beta; \theta) = N\left(R(\theta)\begin{pmatrix}\alpha^2 & 0\\ 0 & \beta^2\end{pmatrix}{}^{\mathrm{T}}R(\theta)\right).$$

We abbreviate $\mathcal{N}_0^2(R(\theta))$ as $\mathcal{N}_0^2(\theta)$. We also set $\Lambda = \{(\lambda, \lambda; \theta) \mid \lambda > 0\}$. Since $(\alpha, \beta; \pi/2 + \theta) = (\beta, \alpha; \theta)$, the expression $(\alpha, \beta; \theta)$ is not a global coordinate system. Even if we consider under modulo $\pi/2$, there is no uniqueness of diagonalizing matrix if α is equal to β.

Throughout this section, we fix $\rho = (\alpha, \beta; 0)$. We regard Gaussian measures $(\alpha, \beta; \theta)$ as ellipsoids: (α, β) specifies the length of the axes with the angle θ of major and minor axes. Let X, Y and Z be matrices defined by

$$X = e_{11} = \frac{1}{\gamma}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\ Y = f_{12} = \frac{1}{\gamma}\begin{pmatrix}0 & -1\\ -1 & 0\end{pmatrix},\ Z = e_{+} = \frac{1}{\gamma}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \tag{4.2}$$

where $\gamma = (\alpha^2 + \beta^2)^{1/2}$. Using this expression, we get

$$\begin{aligned}
\exp_\rho rX &= N(U), & U &= \frac{1}{\gamma^2}\begin{pmatrix}\alpha^2(\gamma + r)^2 & 0\\ 0 & \beta^2(\gamma - r)^2\end{pmatrix},\\
\exp_\rho rY &= N(V), & V &= \frac{1}{\gamma^2}\begin{pmatrix}\alpha^2\gamma^2 + \beta^2 r^2 & \gamma^3 r\\ \gamma^3 r & \alpha^2 r^2 + \beta^2\gamma^2\end{pmatrix},\\
\exp_\rho rZ &= N(W), & W &= \frac{1}{\gamma^2}\begin{pmatrix}\alpha^2(\gamma + r)^2 & 0\\ 0 & \beta^2(\gamma + r)^2\end{pmatrix}.
\end{aligned}$$

We notice that Y changes the axial angle of the ellipsoid, while X and Z do not, see Figure 1.

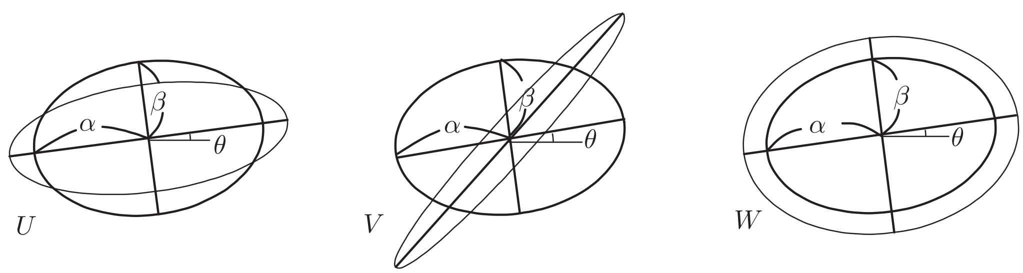

Figure 1

Consequently, $X, Z \in T_\rho\mathcal{N}_0^2(0)$ and $K(X, Z) = 0$ by Corollary 2.5. For changing the axial angle of ellipsoid, the sectional curvature can not vanish.

4.2. *Correspondence to other results*

First, we consider the correspondence to the result of Otto [**13**]. He obtained an explicit expression of sectional curvatures of $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ formally. By making this method rigorous, we give an explicit expression of sectional curvatures of $\mathcal{N}^d$ in [**15**]. He introduced a manifold $\mathcal{M}$ which consists of all diffeomorphisms of $\mathbb{R}^d$ and an isometric submersion from $\mathcal{M}$ into $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ (he also sloppied about a differential structure of $\mathcal{M}$.) He defined a metric g^* on $\mathcal{M}$ which carried the geometry of the L^2-space. Therefore, $(\mathcal{M}, g^*)$ is flat. Using O'Neill's formula [**12**], he showed the sectional curvatures of $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ is given by

$$K(\psi_1, \psi_2)\det\left(g_\rho(\psi_i, \psi_j)\right) = \frac{3}{4}\int_{\mathbb{R}^d}\rho|u|^2 \geq 0,$$

where $\rho \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ and ψ_1, ψ_2, ψ are tangent vectors at ρ given by

$$u = \nabla\psi - [\nabla\psi_1, \nabla\psi_2] \quad \text{and} \quad \mathrm{div}(\rho(\nabla\psi - [\nabla\psi_1, \nabla\psi_2])) = 0.$$

This guarantees that $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ is a space of non-negative curvature. In addition, $K(\psi_1, \psi_2) = 0$ if and only if Hessψ_1 and Hessψ_2 pointwise commute. For X, Y and Z in (4.2), let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be functions whose gradients are X, Y and Z, respectively. Since Z is pointwise commutative with the Hessian of any functions, $K(\mathcal{X}, \mathcal{Z}) = K(\mathcal{Y}, \mathcal{Z}) = 0$ follows. In the case of $\mathcal{X}, \mathcal{Y}$, we demonstrate that Theorem 1.1 coincides with Otto's result.

Let ρ_0 be the standard Gaussian measure on $\mathbb{R}^2$, that is $\rho_0 = (1, 1; \theta)$. Moreover, we define a Gaussian measure ρ and a diffeomorphism Ψ respectively as follows:

$$\rho = (\alpha, \beta; \theta), \quad \Psi(x) = \frac{1}{\gamma} R \begin{pmatrix} \alpha^2 & 0 \\ 0 & \beta^2 \end{pmatrix} {}^{\mathrm{T}}Rx,$$

where $\gamma = (\alpha^2 + \beta^2)^{1/2}$ and $R = R(\theta)$. Then, the submersion sends Ψ to ρ. We choose tangent vectors ψ_1, ψ_2 at ρ as $\mathcal{X}$ and $\mathcal{Y}$, corresponding to X and Y respectively. In terms of Otto's result, we conclude that

$$\begin{aligned}
&g_\rho(\psi_i, \psi_j) = \delta_{ij} \quad (i, j = 1, 2), \\
&[\nabla\psi_1, \nabla\psi_2](x) = \frac{2}{\gamma^2} R \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} {}^{\mathrm{T}}Rx, \\
&\psi(x) = \frac{1}{\gamma^4} {}^{\mathrm{T}}xR \begin{pmatrix} 0 & \alpha^2 - \beta^2 \\ \alpha^2 - \beta^2 & 0 \end{pmatrix} {}^{\mathrm{T}}Rx, \\
&u(x) = \frac{4}{\gamma^4} R \begin{pmatrix} 0 & \alpha^2 \\ -\beta^2 & 0 \end{pmatrix} {}^{\mathrm{T}}Rx.
\end{aligned}$$

Finally, we obtain

$$K(\psi_1, \psi_2) = \frac{3}{4\det(g_\rho(\psi_1, \psi_2))} \int_{\mathbb{R}^2} |u(x)|^2 \rho(x) dx = \frac{12\alpha^2\beta^2}{(\alpha^2 + \beta^2)^3}.$$

Thus we confirm the equivalence between Theorem 1.1 and Otto's result. In [**15**], the sectional curvature of $\mathcal{N}^d$ was also obtained using Riemannian submersion.

4.3. $\mathcal{N}_0^d$ *as Alexandrov spaces.*

Next, we consider the correspondence to results when we regard $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mathcal{N}^d$ as Alexandrov spaces. Details can be found in [**15**].

It is well-known that L^2-Wasserstein space over an Alexandrov space of non-negative curvature is also an Alexandrov space of non-negative curvature (see [**14**, Proposition 2.10].) Therefore $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is an Alexandrov space of non-negative curvature. Since $\mathcal{N}_0^d \subset \mathcal{P}_2(\mathbb{R}^d)$ is a geodesically convex subset, $(\mathcal{N}_0^d, W_2)$ is also an Alexandrov space of non-negative curvature (But it is not complete, the completion of $\mathcal{N}^d$ is given in [**15**].)

Lott and Villani [**8**] made Otto's results rigorous by looking at the space of probability measures as an Alexandrov space, They treated the space of probability measures $\mathcal{P}_2(M)$ over a smooth compact connected manifold M, and proved that M has non-negative sectional curvature if and only if $\mathcal{P}_2(M)$ has non-negative Alexandrov curvature ([**8**, Theorem A.2].) They moreover defined the angle between the geodesics in $\mathcal{P}_2^{\mathrm{ac}}(M)$ ([**8**, Theorem A.17].) We demonstrate it in the case of M as $\mathbb{R}^d$, while $\mathbb{R}^d$ is not a compact manifold. Fix $\rho \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$. If ϕ is a function on $\mathbb{R}^d$ so that $\phi + |\cdot|^2/2$ is convex, then it is regarded as a tangent vector of $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ at ρ. Let ϕ and ψ be such functions. If we set

$$\mu(t) = [\mathrm{id} + t\phi]_\sharp\rho, \ \nu(t) = [\mathrm{id} + t\psi]_\sharp\rho$$

for $t \in [0,1]$, then $\mu(t)$ and $\nu(t)$ are geodesics starting at ρ. The angle between $\mu(t)$ and $\nu(t)$ is given by

$$\cos\measuredangle(\mu,\nu) = \frac{\int_{\mathbb{R}^d}\langle\nabla\phi(x),\nabla\psi(x)\rangle d\rho(x)}{\sqrt{\int_{\mathbb{R}^d}|\nabla\phi(x)|^2 d\rho(x)}\sqrt{\int_{\mathbb{R}^d}|\nabla\psi(x)|^2 d\rho(x)}}.$$

Corresponding to this, we can measure the angle between $\mathcal{N}_0^d(\theta)$ and $\mathcal{N}_0^d(\varphi)$.

PROPOSITION 4.2. *For θ and $\varphi \in (-\pi/4, \pi/4]$, the angle between $\mathcal{N}_0^d(\theta)$ and $\mathcal{N}_0^d(\varphi)$ is $2|\theta-\varphi|$.*

Proof. Let θ and φ be as above. For any $\alpha, \beta, \lambda > 0$,

$$W_2((\alpha,\beta;\theta),(\lambda,\lambda;\theta))^2 = 2\left(\lambda - \frac{\alpha+\beta}{2}\right)^2 + \frac{1}{2}(\alpha-\beta)^2.$$

Hence, the distance from $(\alpha,\beta;\theta)$ to Λ is $(\alpha-\beta)/\sqrt{2}$ and the image of the nearest point projection is

$$\rho = \left(\frac{\alpha+\beta}{2}, \frac{\alpha+\beta}{2}; \theta\right).$$

Let $X(\theta)$ and $X(\varphi)$ be symmetric matrices given by

$$X(\theta) = R(\theta)\begin{pmatrix}(\alpha+\beta)^{-1}(\alpha-\beta) & 0\\ 0 & (\alpha+\beta)^{-1}(\alpha-\beta)\end{pmatrix}R(-\theta)$$
$$X(\varphi) = R(\varphi)\begin{pmatrix}(\alpha+\beta)^{-1}(\alpha-\beta) & 0\\ 0 & (\alpha+\beta)^{-1}(\alpha-\beta)\end{pmatrix}R(-\varphi),$$

then we get

$$\exp_\rho X(\theta) = (\alpha,\beta;\theta), \quad \exp_\rho X(\varphi) = (\alpha,\beta;\varphi).$$

Since

$$g_\rho(X(\theta),X(\varphi)) = \frac{(\alpha+\beta)^2}{4}\mathrm{tr}X(\theta)X(\varphi) = \frac{1}{2}(\alpha-\beta)^2\cos 2(\theta-\varphi)$$

and

$$g_\rho(X(\theta),X(\theta)) = g_\rho(X(\varphi),X(\varphi)) = \frac{1}{2}(\alpha-\beta)^2,$$

we have

$$\frac{g_\rho(X(\theta),X(\varphi))}{\sqrt{g_\rho(X(\theta),X(\theta))g_\rho(X(\varphi),X(\varphi))}} = \cos 2(\theta-\varphi). \tag{4.3}$$

Therefore we conclude

$$\measuredangle(\mathcal{N}_0^d(\theta),\mathcal{N}_0^d(\varphi)) = \mathrm{Arccos}\frac{g_\rho(X(\theta),X(\varphi))}{\sqrt{g_\rho(X(\theta),X(\theta))g_\rho(X(\varphi),X(\varphi))}} = 2|\theta-\varphi|.$$

$\square$

References

1. S. Amari, *Differential-geometrical methods in statistics*, Lecture Notes in Statistics, vol. 28, Springer-Verlag, New York, 1985.
2. Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), no. 4, 375–417.
3. J.A. Carrillo, R.J. McCann, and C. Villani, *Contractions in the 2-Wasserstein length space and thermalization of granular media*, Arch. Ration. Mech. Anal. **179** (2006), no. 2, 217–263.
4. D.C. Dowson and B.V. Landau, *The Fréchet distance between multivariate normal distributions*, J. Multivariate Anal. **12** (1982), no. 3, 450–455.
5. S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry*, third ed., Universitext, Springer-Verlag, Berlin, 2004.
6. W. Gangbo and R.J. McCann, *The geometry of optimal transportation*, Acta Math. **177** (1996), no. 2, 113–161.
7. C.R. Givens and R.M. Shortt, *A class of Wasserstein metrics for probability distributions*, Michigan Math. J. **31** (1984), no. 2, 231–240.
8. J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport compact alexandrov spacess*, Annals of Mathmatics (to appear).
9. R.J. McCann, *A convexity principle for interacting gases*, Adv. Math. **128** (1997), no. 1, 153–179.
10. M.Knott and C.S. Smith, *On the optimal mapping of distributions*, J. Optim. Theory Appl. **43** (1984), no. 1, 39–49.
11. I. Olkin and F.Pukelsheim, *The distance between two random vectors with given dispersion matrices*, Linear Algebra Appl. **48** (1982), 257–263.
12. B. O'Neill, *The fundamental equations of a submersion*, Michigan Math. J. **13** (1966), 459–469.
13. F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.
14. K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math. **196** (2006), no. 1, 65–131.
15. A. Takatsu, *Wasserstein geometry of Gaussian measures*, in preparation.
16. C. Villani, *Optimal transport, new and old*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, Berlin, 2008.

Asuka Takatsu
Mathematical Institute, Tohoku University
Sendai 980-8578, Japan

sa6m21@math.tohoku.ac.jp